

Template-Based Image Reconstruction from Sparse Tomographic Data

Lukas F. Lang¹, Sebastian Neumayer², Ozan Öktem³, and
Carola-Bibiane Schönlieb¹

¹Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom

²Department of Mathematics, Technische Universität Kaiserslautern, D-67663 Kaiserslautern, Germany

³Department of Mathematics, KTH-Royal Institute of Technology, 100 44 Stockholm, Sweden

Abstract

We propose a variational regularisation approach for the problem of template-based image reconstruction from indirect, noisy measurements as given, for instance, in X-ray computed tomography. An image is reconstructed from such measurements by deforming a given template image. The image registration is directly incorporated into the variational regularisation approach in the form of a partial differential equation that models the registration as either mass- or intensity-preserving transport from the template to the unknown reconstruction. We provide theoretical results for the proposed variational regularisation for both cases. In particular, we prove existence of a minimiser, stability with respect to the data, and convergence for vanishing noise when either of the abovementioned equations is imposed and more general distance functions are used. Numerically, we solve the problem by extending existing Lagrangian methods and propose a multilevel approach that is applicable whenever a suitable downsampling procedure for the operator and the measured data can be provided. Finally, we demonstrate the performance of our method for template-based image reconstruction from highly undersampled and noisy Radon transform data. We compare results for mass- and intensity-preserving image registration, various regularisation functionals, and different distance functions. Our results show that very reasonable reconstructions can be obtained when only few measurements are available and demonstrate that the use of a normalised cross correlation-based distance is advantageous when the image intensities between the template and the unknown image differ substantially.

1 Introduction

In medical imaging, an image can typically not be observed directly but only through indirect and potentially noisy measurements, as it is the case, for example, in computed tomography (CT) [41]. Due to the severe ill-posedness of the problem, reconstructing an image from measurements is rendered particularly challenging when only few or partial

measurements are available. This is, for instance, the case in limited-angle computed tomography [22, 41], where limited-angle data is acquired in order to minimise exposure time of organisms to X-radiation. Therefore, it can be beneficial to impose a priori information on the reconstruction, for instance, in the form of a template image. However, typically neither its exact position nor its exact shape is known.

In *image registration*, the goal is to find a reasonable deformation of a given template image so that it matches a given target image as closely as possible according to a predefined similarity measure, see [39, 40] for an introduction. When the target image is unknown and only given through indirect measurements, it is referred to as *indirect image registration* and has been explored only recently [12, 24, 31, 45]. As a result, a deformation together with a transformed template can be computed from tomographic data. The prescribed template acts as a prior for the reconstruction and, when chosen reasonably close in a deformation sense, gives outstanding reconstructions in situations where only few measurements are available and competing methods such as filtered backprojection [41] or total variation regularisation [47] fail, see [12, Sec. 10].

In our setting, deformations are maps from the image domain $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}$, to itself together with an action that specifies exactly how such a map deforms elements in the shape space, which in this work is the space $L^2(\Omega, \mathbb{R})$ of greyscale images supported in the image domain. Natural problems are to characterise admissible deformations and to compute these numerically in an efficient manner.

One possible approach is diffeomorphic image registration, where the set of admissible deformations is restricted to diffeomorphisms in order to preserve the topology of structures within an image [58]. One can, for instance, consider the group of diffeomorphisms together with the composition as group operation. Elements in this group act on greyscale images by means of the group action and thereby allow for a rich set of non-rigid deformations, as required in many applications. For instance, the geometric group action transforms greyscale images in a way such that its intensity values are preserved, whereas the mass-preserving group action ensures that, when the image is regarded as a density, the integral over the density is preserved.

A computational challenge in using the above group formalism is that it lacks a natural vector space structure, which is typically desired for the numerical realisation of the scheme. Hence, it is convenient to further restrict the set of admissible deformations. One way to obtain diffeomorphic deformations is to perturb the identity map with a displacement vector field. Provided that the vector field is reasonably small and sufficiently regular, the resulting map is invertible [58, Prop. 8.6]. For indirect image registration this idea was pursued in [45].

The basic idea of the *large deformation diffeomorphic mapping (LDDMM)* [4, 18, 36, 37, 50, 53, 58] framework is to generate large deformations by considering flows of diffeomorphisms that arise as the solution of an ordinary differential equation (ODE), the so-called flow equation, with velocity fields that stem from a reproducing kernel Hilbert space. In order to ensure that the flow equation admits a unique solution, one typically chooses this vector space so that it can be continuously embedded into $C^1(\Omega, \mathbb{R}^n)$, allowing the application of existence and uniqueness results from Cauchy–Lipschitz theory for ODEs, see [15, Chap. 1] for a brief introduction. In [12], the LDDMM framework is adapted for indirect image registration and the authors prove existence, stability, and convergence of solutions for their variational formulation. Numerically, the problem is solved by gradient descent.

The variational problem associated with LDDMM is typically formulated as an ODE-constrained optimisation problem. As the flow equation can be directly related to hyperbolic partial differential equations (PDE) via the method of characteristics [21, Chap. 3.2], the problem can equivalently be rephrased as a PDE-constrained optimisation problem [33]. The resulting PDE is determined by the chosen group action, see [12, Sec. 6.1.1] for a brief discussion. For instance, the geometric group action is associated with the transport (or advection) equation, while the mass-preserving group action is associated with the continuity equation. It is important to highlight that the PDE constraint implements both the flow equation and the chosen diffeomorphic group action.

Such an optimal control approach was also pursued for motion estimation and image interpolation [2, 5, 6, 9, 13, 29, 44]. In the terminology of optimal control, the PDE represents the state equation, the velocity field the control, and the transformed image the resulting state. We refer to the books [7, 16, 26, 32] and to the article [30] for a general introduction to PDE-constrained optimisation and suitable numerical methods. Let us mention that other methods, such as geodesic shooting [3, 36, 49, 56], exist and constitute particularly efficient numerical approaches. In particular, this direction has recently been combined with machine learning methods [57].

A particularly challenging scenario for diffeomorphic image registration occurs when the target image is not contained in the orbit of the template image under abovementioned group action of diffeomorphisms. For instance, this could happen in the case of the geometric group action due to the appearance of new structures in the target image or due to a discrepancy between the image intensities of the template and the target image. A possible solution is provided by the *metamorphosis* framework [38, 46, 51, 52], which is an extension to LDDMM that allows for modulations of the image intensities along characteristics of the flow. The image intensities change according to an additional flow equation with an unknown source. See [58, Chap. 13] for a general introduction and, for instance, [33] for an application to magnetic resonance imaging. Let us also mention [43], which adopts a discrete geodesic path model for the purpose of image reconstruction, and [34], in which the metamorphosis model is combined with optimal transport.

In [24], the metamorphosis framework is adapted for indirect image registration. The authors prove that their formulation constitutes a well-defined regularisation method by showing existence, stability, and convergence of solutions. However, in the setting where only few measurements—e.g. a few directions in CT—are available, reconstruction of appearing or disappearing structures seems very challenging.

Therefore, in order to obtain robustness with respect to differences in the intensities between the transformed template and the sought target image, we follow a different approach. We consider not only the standard *sum-of-squared differences (SSD)* but also a distance that is based on the *normalised cross correlation (NCC)* [39, Chap. 7.2], as it is invariant with respect to a scaling of the image intensities.

While image registration itself is already an ill-posed inverse problem that requires regularisation [20], the indirect setting as described above is intrinsically more challenging. It can be phrased as an inverse problem, where measurements (or observations) $g \in Y$ are related to an unknown quantity $f \in X$ via the operator equation

$$K(f) = g + n^\delta. \quad (1)$$

Here, $K: X \rightarrow Y$ is a (not necessarily linear) operator that models the data acquisition,

often by means of a physical process, n^δ are measurement errors such as noise, and X and Y are Banach spaces. When f constitutes an image and g are tomographic measurements, solving (1) is often referred to as *image reconstruction*.

We use a variational scheme [48] to solve the inverse problem of indirect image registration, which can be formulated as a PDE-constrained optimisation problem [12, Sec. 6.1.1]. It is given by

$$\begin{aligned} \min_{v \in V} J_{\gamma, g}(v), \\ \text{s.t. } C(v), \end{aligned} \quad (2)$$

where $J_{\gamma, g}: V \rightarrow [0, +\infty]$ is the functional

$$v \mapsto D(K(f_v(T, \cdot)), g) + \gamma \|v\|_V^2. \quad (3)$$

Here, V is an admissible vector space with norm $\|\cdot\|_V$, $D: Y \times Y \rightarrow \mathbb{R}_{\geq 0}$ is a data fidelity term that quantifies the misfit of the solution against the measurements, and $\gamma > 0$ is a regularisation parameter. Moreover, $f_v(T, \cdot): \Omega \rightarrow \mathbb{R}$ denotes the evaluation at time $T > 0$ of the (weak) solution of $C(v)$, which is either the Cauchy problem

$$C(v) = \begin{cases} \frac{\partial}{\partial t} f(t, x) + v(t, x) \nabla_x f(t, x) = 0, & \text{for } (t, x) \in [0, T] \times \Omega, \\ f(0, x) = f_0(x), & \text{for } x \in \Omega, \end{cases}$$

governed by the transport equation, or

$$C(v) = \begin{cases} \frac{\partial}{\partial t} f(t, x) + \operatorname{div}_x(v(t, x) f(t, x)) = 0, & \text{for } (t, x) \in [0, T] \times \Omega, \\ f(0, x) = f_0(x), & \text{for } x \in \Omega, \end{cases}$$

involving the continuity equation. Here, $f_0 \in L^2(\Omega, \mathbb{R})$ denotes an initial condition, which in our case is the template image.

The main goals of this article are the following. First, to study variational and regularising properties of problem (2), and to develop efficient numerical methods for solving it. Second, to investigate alternative choices of distance functions D , such as the abovementioned NCC-based distance. Third, to demonstrate experimentally that excellent reconstructions can be computed from highly undersampled and noisy Radon transform data.

Our numerical approach is based on the Lagrangian methods developed in [35], called LagLDDMM. In contrast to most existing approaches, which are mainly first-order methods (see [35] for a brief classification and discussion), LagLDDMM uses a Gauss–Newton–Krylov method paired with Lagrangian solvers for the hyperbolic PDEs listed above. The characteristics associated with these PDEs are computed with an explicit Runge–Kutta method. One of the main advantages of this approach is that Lagrangian methods are unconditionally stable with regard to the admissible step size. Furthermore, the approach limits numerical diffusion and, in order to evaluate the gradient or the Hessian required for optimisation, does not require the storage of multiple space-time vector fields or images at intermediate time instants. The scheme can also be implemented matrix-free.

In comparison to abovementioned existing methods for indirect image registration, such as [12, 24, 31, 45], our method is conceptually different in several ways. The first difference concerns the discretisation. While [12, 24, 45] are mainly based on small deformations and use reproducing kernel Hilbert spaces, our method relies on nonparametric registration.

The main advantages are that it directly allows for a multilevel approach and no kernel parameters need to be chosen. Moreover, due to the flexibility of the underlying framework it is straightforward to extend our method to parametric registration. Second, our approach relies on second-order methods for optimisation by using a Gauss–Newton method paired with line search, while the other methods mainly rely on gradient descent. This allows for a fast decrease of the objective within only few iterations. Third, our method allows to easily exchange the underlying PDE solver. Essentially, any solver can be used as long as it can be differentiated efficiently. The used explicit Runge–Kutta method has the advantage that it does not require the storage of multiple images or repeated interpolation of the template, which can potentially lead to a blurred solution. Finally, let us mention that [31] is conceptually different since both a deformation and a template image are computed. Our main focus, however, are applications where only very few and noisy measurements are available and the problem of estimating an additional template seems highly underdetermined in such situations.

Contributions The contributions of this article are as follows. First, we provide the necessary theoretical background on (weak) solutions of the continuity and the transport equation, and recapitulate existence and uniqueness theory for characteristic curves for the associated ODE. In contrast to the results derived in [12], where the template image is assumed to be contained in the space $SBV(\Omega, \mathbb{R}) \cap L^\infty(\Omega, \mathbb{R})$ of essentially bounded functions with special bounded variation, our results only require $L^2(\Omega, \mathbb{R})$ regularity. In addition, by using results from [17], we are able to consider the transport equation in the setting with H^1 regularity of vector fields in space as well as in time and with bounded divergence. Moreover, we show the existence of a minimiser of the problem (2), stability with respect to the data, and convergence for vanishing noise.

Second, in order to solve the problem numerically, we follow a *discretise-then-optimize* approach and extend the LagLDDMM framework [35] to the indirect setting. The library itself is an extension of FAIR [39] and, as a result, our implementation provides great flexibility regarding the selected PDE, and can easily be extended to other distances as well as to other regularisation functionals. The source code of our MATLAB implementation is available online.¹

Finally, we present numerical results for the abovementioned distances and PDEs. To the best of our knowledge, the results obtained for indirect image reconstruction based on the continuity equation are entirely novel. Moreover, we propose to use the NCC-based distance instead of SSD whenever the image intensities of the template and the unknown target are far apart, and show its numerical feasibility.

2 Theoretical results on the transport and continuity equation

In this section, we review the necessary theoretical background, and state results on the existence and stability of weak solutions of the transport and the continuity equation. Compared to [12], our results are stronger since we do not require space regularity of the template image.

¹<https://doi.org/10.5281/zenodo.2598138>

2.1 Continuity equation

In what follows, we consider well-posedness of the continuity equation that arises in the LD-MMM framework using the mass-preserving group action via the method of characteristics. The regularity assumptions on v are such that we can apply the theory from [51].

Let $\Omega \subset \mathbb{R}^n$ be a bounded, open, convex domain with Lipschitz boundary and let $T > 0$. In the following, we examine the continuity equation

$$\begin{cases} \frac{\partial}{\partial t} f(t, x) + \operatorname{div}_x(v(t, x)f(t, x)) = 0 & \text{for } (t, x) \in [0, T] \times \Omega, \\ f(0, x) = f_0(x) & \text{for } x \in \Omega, \end{cases} \quad (4)$$

with coefficients $v \in L^2([0, T], \mathcal{V})$ and initial condition $f_0 \in L^2(\Omega, \mathbb{R})$, where \mathcal{V} is a Banach space which is continuously embedded into $C_0^{1,\alpha}(\Omega, \mathbb{R}^n)$ for some $\alpha > 0$. Here $C_0^{1,\alpha}(\Omega, \mathbb{R}^n)$ denotes the closure of $C_c^\infty(\Omega, \mathbb{R}^n)$ under the $C^{1,\alpha}$ norm. Note that such velocity fields can be continuously extended to the boundary. Clearly, equation (4) has to be understood in a weak sense, i.e. a function $f \in C^0([0, T], L^2(\Omega, \mathbb{R}))$ is said to be a weak solution of (4) if

$$\int_0^T \int_\Omega f(t, x) \left(v(t, x) \nabla_x \eta(t, x) + \frac{\partial}{\partial t} \eta(t, x) \right) dx dt + \int_\Omega f_0(x) \eta(0, x) dx = 0 \quad (5)$$

holds for all $\eta \in C_c^\infty([0, T] \times \Omega)$. The corresponding characteristic ODE is

$$\begin{cases} \frac{d}{dt} X(t, s, x) = v(t, X(t, s, x)) & \text{for } (t, s, x) \in [0, T] \times [0, T] \times \Omega, \\ X(s, s, x) = x & \text{for } x \in \Omega. \end{cases} \quad (6)$$

In this notation, the first argument of X is the time dependence, the second the initial time, and the third the initial space coordinate. The following theorem is a reformulation of [51, Thms. 1 and 9] and characterises solutions of (6).

Theorem 2.1.1. *Let $v \in L^2([0, T], \mathcal{V})$ and $s \in [0, T]$ be given. There exists a unique global solution $X(\cdot, s, \cdot) \in C^0([0, T], C^1(\overline{\Omega}, \mathbb{R}^n))$ such that $X(s, s, x) = x$ for all $x \in \Omega$ and*

$$\frac{d}{dt} X(t, s, x) = v(t, X(t, s, x))$$

in weak sense (absolutely continuous solutions). The solution operator $X_v: L^2([0, T], \mathcal{V}) \rightarrow C^0([0, T] \times \overline{\Omega}, \mathbb{R}^n)$ assigning a flow X_v to every velocity field v is continuous with respect to the weak topology in $L^2([0, T], \mathcal{V})$.

Since $X(0, t, X(t, 0, x)) = x$, we can directly conclude that $X(t, 0, \cdot)$ is a diffeomorphism for every $t \in [0, T]$. Now, the diffeomorphism $X(0, t, x)$ can be used to characterise solutions of (4) as follows.

Proposition 2.1.2. *If $v \in L^2([0, T], \mathcal{V})$, then the unique weak solution of (4), as defined in (5), is given by $f(t, x) = \det(\mathcal{D}_x X(0, t, x)) f_0(X(0, t, x))$, where $\mathcal{D}_x X$ denotes the Jacobian of X .*

Proof. The proof is divided in three steps. First, we want to show that f satisfies the regularity conditions of weak solutions. For this purpose, the first step is to show $X(0, \cdot, \cdot) \in C^0([0, T], C^0(\overline{\Omega}, \mathbb{R}^n))$, i.e. that the flow is continuous in the initial values. Clearly, $X(0, t, \cdot) \in C^0(\overline{\Omega}, \mathbb{R}^n)$ for every $t \in [0, T]$. For an arbitrary sequence $t_i \rightarrow t$ we get

$$\|X(0, t_i, \cdot) - X(0, t, \cdot)\|_{C^0(\overline{\Omega})} \leq \|\mathcal{D}_x X(0, t_i, \cdot)\|_{C^0(\overline{\Omega})} \|\text{Id} - X(t_i, t, \cdot)\|_{C^0(\overline{\Omega})} \rightarrow 0,$$

where the first factor is bounded due to [51, Lemma 9]. Next, using the sequence $X_i(\cdot) = X(0, t_i, \cdot)$, it follows $f_0(X(0, \cdot, \cdot)) \in C^0([0, T], L^2(\Omega, \mathbb{R}))$, where the continuity in time follows from [42, Cor. 3]. Then, by differentiating $X(0, t, X(t, 0, x)) = x$ and rearranging the terms we obtain

$$\det(\mathcal{D}_x X(0, \cdot, \cdot)) = \det(\mathcal{D}_x X(\cdot, 0, \cdot))^{-1}(X(0, \cdot, \cdot)) \in C^0([0, T] \times \overline{\Omega}),$$

since all involved expressions are continuous. Finally, we conclude $f \in C^0([0, T], L^2(\Omega, \mathbb{R}))$, which follows from

$$\begin{aligned} & \|f(t, \cdot) - f(t_i, \cdot)\|_{L^2(\Omega)} \\ & \leq \|\det(\mathcal{D}_x X(0, t, x)) - \det(\mathcal{D}_x X(0, t_i, x))\|_{C^0(\overline{\Omega})} \|f_0(X(0, t, x))\|_{L^2(\Omega)} \\ & \quad + \|\det(\mathcal{D}_x X(0, t_i, x))\|_{C^0(\overline{\Omega})} \|f_0(X(0, t, x)) - f_0(X(0, t_i, x))\|_{L^2(\Omega)}, \end{aligned} \quad (7)$$

since both summands converge to zero.

The second step is to show that (5) is satisfied. Note that $X(\cdot, 0, x)$ is differentiable in t for a.e. $t \in [0, T]$, since it is absolutely continuous by definition. By inserting f into (5) and using the transformation formula, we get

$$\begin{aligned} & \int_0^T \int_{\Omega} f(t, x) \left(v(t, x) \nabla_x \eta(t, x) + \frac{\partial \eta(t, x)}{\partial t} \right) dx dt + \int_{\Omega} f_0(x) \eta(0, x) dx \\ & = \int_0^T \int_{\Omega} \det(\mathcal{D}_x X(t, 0, x)) f(t, X(t, 0, x)) \frac{d}{dt} \eta(t, X(t, 0, x)) dx dt + \int_{\Omega} f_0(x) \eta(0, x) dx \\ & = \int_0^T \int_{\Omega} f_0(x) \frac{d}{dt} \eta(t, X(t, 0, x)) dx dt + \int_{\Omega} f_0(x) \eta(0, x) dx = 0. \end{aligned} \quad (8)$$

For the last equality we used that $\eta(t, X(t, 0, x))$ is absolutely continuous.

The last step is to prove uniqueness of weak solutions, i.e. that every solution has the given form. Let f_1, f_2 be two different solutions, then we can find a t such that $\|f_1(t, \cdot) - f_2(t, \cdot)\|_{L^2(\Omega)} > 0$. By continuity in time we can find an interval I of length $\delta > 0$ that contains t , and a constant $c > 0$ such that

$$\|f_1(s, \cdot) - f_2(s, \cdot)\|_{L^2(\Omega)} \geq c$$

for all $s \in I$. However, weak solutions are unique in $L^\infty([0, T], L^2(\Omega, \mathbb{R}))$, see [17, Cor. II.1], where we used the embedding of \mathcal{V} into $C_0^1(\Omega, \mathbb{R}^n)$. This yields a contradiction. \square

Additionally, we can state and prove the following stability result for solutions of (4).

Proposition 2.1.3 (Stability). *Let $v_i \rightharpoonup v$ in $L^2([0, T], \mathcal{V})$ and f_i denote the weak solution of (4) corresponding to v_i . Then for every $t \in [0, T]$, there exists a subsequence, also denoted with f_i , such that $f_i(t, \cdot) \rightarrow f(t, \cdot)$ in $L^2(\Omega, \mathbb{R})$.*

Proof. The solution of (6) corresponding to v_i is denoted by X_i . Fix an arbitrary $t \in [0, T]$. From Thm. 2.1.1 we conclude $\|X_i(0, t, \cdot) - X(0, t, \cdot)\|_{C^0(\bar{\Omega})} \rightarrow 0$. Further, [19, Thm. 3.1.10] implies that $X_i(0, t, \cdot)$ is uniformly bounded for all $i \in \mathbb{N}$ in $C^{1,\alpha}(\bar{\Omega})$, which implies $f_0(X_i(0, t, \cdot)) \rightarrow f_0(X(0, t, \cdot))$ in $L^2(\Omega, \mathbb{R})$ by [42, Cor. 3].

It is left to show that a subsequence, also denoted by X_i , exists such that $X_i(0, t, \cdot) \rightarrow X(0, t, \cdot)$ in $C^1(\bar{\Omega}, \mathbb{R}^n)$. This concludes the proof since it also implies the convergence of $\det(\mathcal{D}_x X_i(0, t, \cdot)) \rightarrow \det(\mathcal{D}_x X(0, t, \cdot))$ in $C^0(\bar{\Omega})$. However, $X_i(0, t, \cdot)$ is uniformly bounded in $C^{1,\alpha}(\bar{\Omega}, \mathbb{R}^n)$ and it follows that $\mathcal{D}_x X_i(0, t, \cdot)$ is uniformly bounded in $C^{0,\alpha}(\bar{\Omega}, \mathbb{R}^{n \times n})$. By using the compact embedding of $C^{0,\alpha}(\bar{\Omega}, \mathbb{R}^{n \times n})$ into $C^0(\bar{\Omega}, \mathbb{R}^{n \times n})$ [23, Lemma 6.33], there exists a subsequence of $X_i(0, t, \cdot)$ that converges to $X(0, t, \cdot)$ in $C^1(\bar{\Omega}, \mathbb{R}^n)$. \square

2.2 Transport equation with H^1 regularity

Here, we prove well-posedness of the transport equation that arises in the LDDMM framework using the geometric group action. Compared to the previous section, the space regularity assumptions on v are weaker and fit the setting in [17].

The transport equation reads as

$$\begin{cases} \frac{\partial}{\partial t} f(t, x) + v(t, x) \nabla_x f(t, x) = 0 & \text{for } (t, x) \in [0, T] \times \Omega, \\ f(0, x) = f_0(x) & \text{for } x \in \Omega, \end{cases} \quad (9)$$

with coefficients

$$v \in A := \left\{ v \in H^1([0, T] \times \Omega)^n \cap L^2([0, T], H_0^1(\Omega)^n) : \|\operatorname{div}_x v\|_{L^\infty([0, T] \times \Omega)} \leq C \right\} \quad (10)$$

for some fixed constant C and initial value $f_0 \in L^2(\Omega, \mathbb{R})$. The admissible set A consists of all H^1 functions that are zero on the boundary of the spatial domain and have bounded divergence in the L^∞ norm.

Note that the set A is a subset of $H^1([0, T] \times \Omega)^n$, which is closed and convex so that it is a weakly closed subset of a reflexive Banach space. In the following, we only check that A is closed. Let v_i be a convergent sequence in A with limit v . Since the two involved spaces are Banach spaces, we only have to check that v satisfies the constraint. Assume that $\|\operatorname{div}_x v\|_{L^\infty([0, T] \times \Omega)} > C$, then there exists a set B with positive measure and an $\epsilon > 0$ such that for all $x \in B$ we have $|\operatorname{div}_x v(x)| \geq C + \epsilon$. Hence, we get $\|\operatorname{div}_x v_i - \operatorname{div}_x v\|_{L^2([0, T] \times \Omega)} \geq \sqrt{\mu(B)}\epsilon$, which contradicts the convergence in H^1 .

Again, equation (9) has to be understood in the weak sense so that $f \in C^0([0, T], L^2(\Omega, \mathbb{R}))$ is said to be a solution of (9) if it satisfies

$$\int_0^T \int_\Omega f(t, x) \left(\operatorname{div}_x (v(t, x) \eta(t, x)) + \frac{\partial}{\partial t} \eta(t, x) \right) dx dt + \int_\Omega f_0(x) \eta(0, x) dx = 0 \quad (11)$$

for all $\eta \in C_c^\infty([0, T] \times \Omega)$. The next theorem is an existence and stability result, see [17, Cors. II.1 and II.2, Thm. II.5].

Theorem 2.2.1 (Existence & Stability). *For every $v \in A$ there exists a unique weak solution $f \in C^0([0, T], L^2(\Omega, \mathbb{R}))$ of (9). If $v_i \in A$ converges to $v \in A$ in the norm of $L^2([0, T] \times \Omega, \mathbb{R}^n)$, then the corresponding sequence of weak solutions $f_i \in C^0([0, T], L^2(\Omega, \mathbb{R}))$ converges to f in $C^0([0, T], L^2(\Omega, \mathbb{R}))$.*

Proof. The existence and uniqueness of weak solutions follows from [17, Corrs. II.1 and II.2]. Note that these solutions are also renormalised due to [17, Thm. II.3].

We recast the second part of the theorem such that it has the exact form of [17, Thm. II.5]. First, note that both the velocity fields and the initial condition can be extended to \mathbb{R}^n by zero outside of Ω due to boundary condition of A . Due to the conditions on v , the weak formulation is equivalent to the one for the extension in the \mathbb{R}^n setting. The uniform boundedness condition on f_i is satisfied since Ω is bounded. \square

Corollary 2.2.2. *Let $v_i \rightharpoonup v \in A$ with the inner product of $H^1([0, T] \times \Omega)^n$. Then, f_i converges to f in $C^0([0, T], L^2(\Omega, \mathbb{R}))$.*

Proof. Combine the previous theorem with the compact embedding of $H^1([0, T] \times \Omega)^n$ into $L^2([0, T] \times \Omega)^n$ (Rellich embedding theorem [1, A6.4]). \square

Remark 2.2.3. *Note that the same arguments can be used if we use higher spatial regularity, such as H^2 , in this section. From a numerical point of view, the bound on the divergence is always satisfied for C large enough if we use linear interpolation for the velocities on a fixed grid. Here we use that all norms are equivalent on finite dimensional spaces.*

3 Regularising properties of template-based image reconstruction

In this section, we prove regularising properties of template-based reconstruction as defined in (2). Recall that the problem reads

$$\begin{aligned} \min_{v \in V} D(K(f_v(T, \cdot)), g) + \gamma \|v\|_V^2, \\ \text{s.t. } C(v), \end{aligned}$$

where $C(v)$ is the Cauchy problem with either the transport or the continuity equation. The admissible set V is chosen such that the regularity requirements stated in the previous section are satisfied. For the following considerations we require these assumptions on K and D :

1. The operator K is continuous, $D(\cdot, g)$ is lower semi-continuous for each $g \in Y$, and $D(g, \cdot)$ is continuous for each $g \in Y$.
2. If f_n, g_n are two convergent sequences with limits f and g , respectively, then D must satisfy $\liminf_{n \rightarrow \infty} D(f_n, g) \leq \liminf_{n \rightarrow \infty} D(f_n, g_n)$.
3. If $D(f, g) = 0$, then $f = g$.

Note that the requirements on D are satisfied if D is a metric. The obtained results are along the lines of [12] but are adapted to our setting and to our notation. For simplicity we stick to the notation of the continuity equation, but want to mention that the same derivations hold for the transport equation with coefficients in the set A . First, we prove that a minimiser of the problem exists.

Proposition 3.0.1 (Existence). *For every $f_0 \in L^2(\Omega, \mathbb{R})$, the functional $J_{\gamma, g}$ defined in (3) has a minimiser.*

Proof. The idea of the proof is to construct a minimising sequence which is weakly convergent and then use that the functional is weakly lower semi-continuous. Let us consider a sequence v_n such that $J_{\gamma,g}(v_n)$ converges to $\inf_v J_{\gamma,g}(v)$. By construction of the functional, v_n is bounded in $L^2([0, T], \mathcal{V})$ and hence there exists a subsequence, also denoted with v_n , such that $v_n \rightharpoonup v_\infty$. By Prop. 2.1.3, there exists a subsequence, also denoted with v_n , such that $f_{v_n}(T, \cdot) \rightarrow f_{v_\infty}(T, \cdot)$ in $L^2(\Omega, \mathbb{R})$. With this at hand, we are able to prove weak lower semi-continuity of the data term. Indeed, as K is continuous, from $f_{v_n}(T, \cdot) \rightarrow f_{v_\infty}(T, \cdot)$ we get $K(f_{v_n}(T, \cdot)) \rightarrow K(f_{v_\infty}(T, \cdot))$. Since $D(\cdot, g)$ is lower semi-continuous, we obtain that $D(K(f_{v_\infty}(T, \cdot)), g) \leq \liminf_{n \rightarrow \infty} D(K(f_{v_n}(T, \cdot)), g)$. This concludes the proof, since the whole functional is (weakly) lower semi-continuous, and hence $J_{\gamma,g}(v_\infty) \leq \inf_v J_{\gamma,g}(v)$. \square

Next, we state a stability result.

Proposition 3.0.2 (Stability). *Let $f_0 \in L^2(\Omega, \mathbb{R})$ and $\gamma > 0$. Let g_n be a sequence in Y converging to $g \in Y$. For each n , we choose v_n as minimiser of J_{γ,g_n} . Then, there exists a subsequence of v_n which weakly converges towards a minimiser v of $J_{\gamma,g}$.*

Proof. By the properties of D it holds, for every n , that

$$\|v_n\|_V^2 \leq \frac{1}{\gamma} J_{\gamma,g_n}(v_n) \leq \frac{1}{\gamma} J_{\gamma,g_n}(0) = \frac{1}{\gamma} D(K(f_0), g_n) \rightarrow \frac{1}{\gamma} D(K(f_0), g) < \infty.$$

Hence, v_n is bounded in $L^2([0, T], \mathcal{V})$ and there exists a subsequence, also denoted with v_n , such that $v_n \rightharpoonup v$. From the weak convergence we obtain $\gamma\|v\|_V^2 \leq \gamma \liminf_{n \rightarrow \infty} \|v_n\|_V^2$.

By passing to a subsequence and by using Prop. 2.1.3, we deduce that $f_{v_n}(T, \cdot) \rightarrow f_v(T, \cdot)$. Together with the convergence of g_n and the convergence property of D this implies

$$D(K(f_v(T, \cdot)), g) \leq \liminf_{n \rightarrow \infty} D(K(f_{v_n}(T, \cdot)), g) \leq \liminf_{n \rightarrow \infty} D(K(f_{v_n}(T, \cdot)), g_n).$$

Thus, for any \tilde{v} , it holds that

$$J_{\gamma,g}(v) \leq \liminf_{n \rightarrow \infty} \gamma\|v_n\|_V^2 + D(K(f_{v_n}(T, \cdot)), g_n) = \liminf_{n \rightarrow \infty} J_{\gamma,g_n}(v_n) \leq \liminf_{n \rightarrow \infty} J_{\gamma,g_n}(\tilde{v}),$$

because v_n minimises J_{γ,g_n} . Then, as $J_{\gamma,g_n}(\tilde{v})$ converges to $J_{\gamma,g}(\tilde{v})$ by the assumptions on D , we deduce $J_{\gamma,g}(v) \leq J_{\gamma,g}(\tilde{v})$ and hence that v minimises $J_{\gamma,g}$. \square

Finally, we state a convergence result for the method.

Proposition 3.0.3 (Convergence). *Let $f_0 \in L^2(\Omega, \mathbb{R})$ and $g \in Y$, and suppose that there exists $\hat{v} \in L^2([0, T], \mathcal{V})$ such that $K(f_{\hat{v}}(T, \cdot)) = g$. Further, assume that $\gamma : \mathbb{R}_{>0} \mapsto \mathbb{R}_{>0}$ satisfies $\gamma(\delta) \rightarrow 0$ and $\frac{\delta}{\gamma(\delta)} \rightarrow 0$ as $\delta \rightarrow 0$. Now let δ_n be a sequence of positive numbers converging to 0 and assume that g_n is a data sequence satisfying $D(g, g_n) \leq \delta_n$ for each n . Let v_n be a minimiser of J_{γ_n, g_n} , where $\gamma_n = \gamma(\delta_n)$. Then, there exists a subsequence of v_n which weakly converges towards an element v such that $K(f_v(T, \cdot)) = g$.*

Proof. For every n , it holds that

$$\|v_n\|_V^2 \leq \frac{1}{\gamma_n} J_{\gamma_n, g_n}(v_n) \leq \frac{1}{\gamma_n} J_{\gamma_n, g_n}(\hat{v}) = \frac{1}{\gamma_n} (D(g, g_n) + \gamma_n \|\hat{v}\|_V^2) \leq \frac{\delta_n}{\gamma_n} + \|\hat{v}\|_V^2. \quad (12)$$

From the requirements on γ and δ we deduce that v_n is bounded in $L^2([0, T], \mathcal{V})$ and then that up to an extraction, v_n weakly converges to some v in $L^2([0, T], \mathcal{V})$.

Further, it holds $D(K(f_v(T, \cdot)), g) \leq \liminf_{n \rightarrow \infty} D(K(f_{v_n}(T, \cdot)), g_n)$ with the same arguments as in the previous proposition. Finally, for every n , it holds that

$$D(K(f_{v_n}(T, \cdot)), g_n) \leq J_{\gamma_n, g_n}(v_n) \leq J_{\gamma_n, g_n}(\hat{v}) = D(g, g_n) + \gamma_n \|\hat{v}\|_V^2, \quad (13)$$

where the two rightmost terms both converge to zero. Thus, $K(f_v(T, \cdot)) = g$ by the assumptions on D . \square

We conclude with a remark on data discrepancy functionals that satisfy the conditions and will be used in our numerical experiments in Sec. 5.

Remark 3.0.4. *We now assume that the data space Y is a real Hilbert space. Clearly, the conditions are satisfied if $D_{\text{SSD}}(f, g) = \|f - g\|_Y^2$. We will only check the convergence condition. It holds*

$$\liminf_{n \rightarrow \infty} \|f_n - g\|_Y^2 = \liminf_{n \rightarrow \infty} \|f_n - g_n\|_Y^2 + 2\langle f_n - g_n, g_n - g \rangle + \|g - g_n\|_Y^2,$$

where the last two terms converge to zero since convergent sequences are bounded.

Another function that satisfies the conditions is $D_{\text{NCC}}: Y \setminus \{0\} \times Y \setminus \{0\} \rightarrow [0, 1]$ with

$$D_{\text{NCC}}(f, g) = 1 - \frac{\langle f, g \rangle^2}{\|f\|_Y^2 \|g\|_Y^2},$$

which is based on the normalised cross correlation. First, note that $\tilde{D}(\cdot, g) = \frac{\langle \cdot, g \rangle^2}{\|g\|_Y^2}$ and the function $\|\cdot\|_Y^{-2}$ are continuous. Thus, we get that $D_{\text{NCC}}(\cdot, g)$ is continuous. By symmetry, this also holds for $D_{\text{NCC}}(g, \cdot)$. It remains to check the convergence property:

$$\lim_{n \rightarrow \infty} 1 - D_{\text{NCC}}(f_n, g) = \lim_{n \rightarrow \infty} \frac{(\langle f_n, g - g_n \rangle + \langle f_n, g_n \rangle)^2}{\|f_n\|_Y^2 \|g\|_Y^2} = \lim_{n \rightarrow \infty} \frac{\langle f_n, g_n \rangle^2}{\|f_n\|_Y^2 \|g\|_Y^2} \quad (14)$$

$$= \lim_{n \rightarrow \infty} \frac{\langle f_n, g_n \rangle^2}{\|f_n\|_Y^2 \|g_n\|_Y^2} = \lim_{n \rightarrow \infty} 1 - D_{\text{NCC}}(f_n, g_n). \quad (15)$$

From this we conclude $\liminf_{n \rightarrow \infty} D_{\text{NCC}}(f_n, g) = \liminf_{n \rightarrow \infty} D_{\text{NCC}}(f_n, g_n)$. Unfortunately, $D_{\text{NCC}}(f, g) = 0$ only implies $f = cg$, with $c \in \mathbb{R}$.

4 Numerical solution

The focus of this section is to approximately solve problem (2). Our approach is based on the Lagrangian methods developed in [35] and the inexact multilevel Gauss–Newton method used in [39]. Both methods and their necessary modifications are briefly outlined here.

As customary in PDE-constrained optimisation [16, Chap. 3], we eliminate the state equation by defining a *control-to-state* operator, which parametrises the final state $f_v(T, \cdot)$ in terms of the unknown velocities v . With a slight abuse of notation, we define this solution map as

$$\begin{aligned} S: V &\rightarrow L^2(\Omega, \mathbb{R}), \\ v &\mapsto f_v(T, \cdot) =: f(v). \end{aligned} \quad (16)$$

Here, f_v denotes the unique solution to either the transport or the continuity equation, as defined in Sec. 2. As a result, we obtain the reduced form of (2):

$$\min_{v \in V} D(K(f(v)), g) + \gamma R(v). \quad (17)$$

Here, $R: V \rightarrow \mathbb{R}_{\geq 0}$ is a regularisation functional that can be written as

$$R(v) = \frac{1}{2} \int_0^T \int_{\Omega} \|Bv(t, x)\|^2 dx dt \quad (18)$$

with B denoting a linear (vectorial) differential operator.

In this work, we consider the operators $B = \nabla_x$ and $B = \Delta_x$, which are also used in [35]. We refer to the resulting functionals R as *diffusion* and *curvature regularisation* functionals, respectively. Note that B can as well be chosen to incorporate derivatives with respect to time.

Amongst the operators above, we also consider a regularisation functional that resembles the norm of the space $V = L^2([0, T], H_0^3(\Omega, \mathbb{R}^n))$. This particular choice is motivated by the fact that, for $n = \{2, 3\}$, the space $H_0^3(\Omega, \mathbb{R}^n)$ can be continuously embedded in $C_0^{1,\alpha}(\Omega, \mathbb{R}^n)$, for some $\alpha > 0$, so that the results in Sec. 2 hold. The norm of V is given by

$$\|v\|_V^2 = \frac{1}{2} \int_0^T \|v(t, \cdot)\|_{L^2(\Omega, \mathbb{R}^n)}^2 dt + \frac{1}{2} \int_0^T |v(t, \cdot)|_{H^3(\Omega, \mathbb{R}^n)}^2 dt. \quad (19)$$

Here, $|\cdot|_{H^k(\Omega, \mathbb{R}^n)}$ denotes the usual H^k -seminorm including only the highest-order partial derivatives. By the Gagliardo–Nirenberg inequality, (19) is equivalent to the usual norm of $L^2([0, T], H_0^3(\Omega, \mathbb{R}^n))$. To simplify numerical optimisation, we omit the requirement that v is compactly supported in Ω and minimise over $L^2([0, T], H^3(\Omega, \mathbb{R}^n))$.

In order to solve problem (17), we follow a discretise-then-optimize strategy. Without loss of generality, we assume that the domain is $\Omega = (0, 1)^n$. We partition it into a regular grid consisting of m^n equally sized cells of edge length $h_X = 1/m$ in every coordinate direction.

The template image $f_0 \in L^2(\Omega, \mathbb{R})$ is assumed to be sampled at cell-centred locations $\mathbf{x}_c \in \mathbb{R}^{m^n}$, giving rise to its discrete version $\mathbf{f}_0(\mathbf{x}_c) \in \mathbb{R}^{m^n}$. The template image is interpolated on the cell-centred grid by means of cubic B-spline interpolation as outlined in [39, Chap. 3.4].

Similarly, the time domain is assumed to be $[0, 1]$ and is partitioned into m_t equally sized cells of length $h_t = 1/m_t$. We assume that the unknown velocities $v: [0, 1] \times \Omega \rightarrow \mathbb{R}^n$ are sampled at cell-centred locations in space as well as at cell-centred locations in time, leading to a vector of unknowns $\mathbf{v} \in \mathbb{R}^N$, where $N = (m_t + 1) \cdot n \cdot m^n$ is the total number of unknowns of the finite-dimensional minimisation problem.

Lagrangian solver In order to compute the solution map $f(v)$ numerically, i.e. to solve the hyperbolic PDEs (4) and (9), the Lagrangian solver in [35] follows a two-step approach. First, given a vector $\mathbf{v} \in \mathbb{R}^N$ of velocities, the ODE (6) is solved approximately using a fourth-order Runge–Kutta (RK4) method with N_t equally spaced time steps of size Δt . For simplicity, we follow the presentation in [35] based on an explicit first-order Euler method and refer to [35, Sec. 3.1] for the full details.

Given initial points $\mathbf{x} \in \mathbb{R}^{m^n}$ and velocities $\mathbf{v} \in \mathbb{R}^N$, an approximation $\mathbf{X}_{\mathbf{v}}: [0, 1]^2 \times \mathbb{R}^{m^n} \rightarrow \mathbb{R}^{m^n}$ of the solution X_v is given recursively by

$$\mathbf{X}_{\mathbf{v}}(0, t_{k+1}, \mathbf{x}) = \mathbf{X}_{\mathbf{v}}(0, t_k, \mathbf{x}) + \Delta t \mathbf{I}(\mathbf{v}, t_k, \mathbf{X}_{\mathbf{v}}(0, t_k, \mathbf{x})), \quad (20)$$

for all $k = 0, 1, \dots, N_t - 1$, with initial condition $\mathbf{X}_{\mathbf{v}}(0, 0, \mathbf{x}) = \mathbf{x}$. Here, $\mathbf{I}(\mathbf{v}, t_k, \mathbf{X}_{\mathbf{v}}(0, t_k, \mathbf{x}))$ denotes a componentwise interpolation of \mathbf{v} at time $t_k = k\Delta t$ and at the points $\mathbf{X}_{\mathbf{v}}(0, t_k, \mathbf{x})$. Note that, since the characteristic curves for both PDEs coincide, this step is identical regardless of which PDE we impose.

The second step computes approximate intensities of the final state $f_v(1, \cdot)$. This step depends on the particular PDE. For the transport equation, in order to compute the intensities at the grid points \mathbf{x}_c , we follow characteristic curves backwards in time, which is achieved by setting $\Delta t = -1/N_t$ in (20). The deformed template is then given by

$$\mathbf{f}(\mathbf{v}) = \mathbf{f}_0(\mathbf{X}_{\mathbf{v}}(1, 0, \mathbf{x}_c)), \quad (21)$$

where $\mathbf{f}_0 \in \mathbb{R}^{m^n}$ is the interpolation of the discrete template image.

For the continuity equation, [35] proposes to use a *particle-in-cell (PIC)* method, see [14] for details. The density of particles which are initially located at grid points \mathbf{x}_c is represented by a linear combination of basis functions, which are then shifted by following the characteristics computed in the first step. To determine the final density at grid points, exact integration over the grid cells is performed. By setting $\Delta t = 1/N_t$ in (20), the transformed template can be computed as

$$\mathbf{f}(\mathbf{v}) = \mathbf{F}(\mathbf{X}_{\mathbf{v}}(0, 1, \mathbf{x}_c))\mathbf{f}_0(\mathbf{x}_c), \quad (22)$$

where $\mathbf{F} \in \mathbb{R}^{N \times N}$ is the pushforward matrix that computes the integrals over the shifted basis functions. See [35, Sec. 3.1] for its detailed specification using linear, compactly supported basis functions. By design, the method is mass-preserving at the discrete level.

Numerical optimisation Let us denote by $\mathbf{K}: \mathbb{R}^N \rightarrow \mathbb{R}^M$, $M \in \mathbb{N}$, a finite-dimensional, Fréchet differentiable approximation of the (not necessarily linear) operator $K: L^2(\Omega, \mathbb{R}) \rightarrow Y$. With the application to CT in mind, we will outline a discretisation of (17) suitable for the n -dimensional Radon transform, which maps a function on \mathbb{R}^n into the set of its integrals over the hyperplanes in \mathbb{R}^n [41, Chap. 2].

An element $K(f(v)) \in Y$ is a function on the unit cylinder $S^{n-1} \times \mathbb{R}$ of \mathbb{R}^{n+1} , where S^{n-1} is the $(n-1)$ -dimensional unit sphere. We discretise this unit cylinder as follows. First, we sample $p \in \mathbb{N}$ directions from S^{n-1} . When $n = 2$, as it is the case in our experiments in Sec. 5, directions are parametrised by angles from the interval $[0, 180]$ degrees. For simplicity, we say (slightly imprecise) that we take one measurement in each direction. Second, similarly to the sampling of Ω , we use an interval $(0, 1)$ instead of \mathbb{R} and partition it into q equally sized cells of length $h_Y = 1/q$. Depending on n and the diameter of Ω , the interval length may require adjustment. Each measurement i is then sampled at cell-centred points $\mathbf{y}_c \in \mathbb{R}^q$ and denoted by $\mathbf{g}_i(\mathbf{y}_c) \in \mathbb{R}^q$. All measurements are then concatenated into a vector $\mathbf{g} := \mathbf{g}(\mathbf{y}_c) \in \mathbb{R}^M$, where $M = p \cdot q$.

The finite-dimensional optimisation problem in abstract form is then given by

$$\min_{\mathbf{v} \in \mathbb{R}^N} \{J_{\gamma, \mathbf{g}}(\mathbf{v}) := D(\mathbf{K}(\mathbf{f}(\mathbf{v})), \mathbf{g}) + \gamma R(\mathbf{v})\}, \quad (23)$$

where D and R are chosen to be discretisations of a distance and of (18), respectively.

In further consequence, we approximate integrals using a midpoint quadrature rule. As we are mainly interested in the setting where only few directions are given, we disregard

integration over the unit sphere. For vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^M$, the corresponding approximations of the distance based on sum-of-squared-differences and the normalised cross correlation-based distance are then

$$D_{\text{SSD}}(\mathbf{x}, \mathbf{y}) \approx \frac{h_Y}{2}(\mathbf{x} - \mathbf{y})^\top (\mathbf{x} - \mathbf{y}) \quad \text{and} \quad D_{\text{NCC}}(\mathbf{x}, \mathbf{y}) \approx 1 - \frac{(\mathbf{x}^\top \mathbf{y})^2}{\|\mathbf{x}\|^2 \|\mathbf{y}\|^2}, \quad (24)$$

respectively. See [39, Chaps. 6.2 and 7.2] for details. Note that, due to cancellation, no (spatial) discretisation parameter occurs in the approximation of the NCC above.

Moreover, we approximate the regularisation functional in (18) with

$$R(\mathbf{v}) \approx \frac{h_t h_X^n}{2} \mathbf{v}^\top \mathbf{B}^\top \mathbf{B} \mathbf{v}, \quad (25)$$

where $\mathbf{B} \in \mathbb{R}^{N \times N}$ is a finite-difference discretisation of the differential operator in (18), analogous to [40, Chap. 8.5]. In our implementation we use zero Neumann boundary conditions and pad the spatial domain to mitigate boundary effects arising from the discretisation of the operator.

In order to apply (inexact) Gauss–Newton optimisation to problem (23), we require first- and (approximate) second-order derivatives of $J_{\gamma, \mathbf{g}}(\mathbf{v})$. By application of the chain rule, we obtain

$$\frac{\partial}{\partial \mathbf{v}} J_{\gamma, \mathbf{g}}(\mathbf{v}) = \frac{\partial}{\partial \mathbf{v}} \mathbf{f}(\mathbf{v})^\top \frac{\partial}{\partial \mathbf{f}} \mathbf{K}(\mathbf{f}(\mathbf{v}))^\top \frac{\partial}{\partial \mathbf{x}} D(\mathbf{K}(\mathbf{f}(\mathbf{v})), \mathbf{g}) + \gamma \frac{\partial}{\partial \mathbf{v}} R(\mathbf{v}),$$

where $\partial \mathbf{K} / \partial \mathbf{f}$ is the Fréchet derivative of \mathbf{K} and $\partial \mathbf{f}(\mathbf{v}) / \partial \mathbf{v}$ is the derivative of the solution map (16) with respect to the velocities, which is given below.

The partial derivatives of the distance functions (24) with respect to its first argument are given by

$$\frac{\partial}{\partial \mathbf{x}} D_{\text{SSD}}(\mathbf{x}, \mathbf{y}) = h_Y (\mathbf{x} - \mathbf{y}) \quad \text{and} \quad \frac{\partial^2}{\partial \mathbf{x}^2} D_{\text{SSD}}(\mathbf{x}, \mathbf{y}) = h_Y \mathbf{I}_N, \quad (26)$$

where $\mathbf{I}_N \in \mathbb{R}^{N \times N}$ is the identity matrix of size N , and

$$\frac{\partial}{\partial \mathbf{x}} D_{\text{NCC}}(\mathbf{x}, \mathbf{y}) = -\frac{2(\mathbf{x}^\top \mathbf{y})\mathbf{y}}{\|\mathbf{x}\|^2 \|\mathbf{y}\|^2} + \frac{2(\mathbf{x}^\top \mathbf{y})^2 \mathbf{x}}{\|\mathbf{x}\|^4 \|\mathbf{y}\|^2}$$

respectively. Moreover, the derivatives of (25) are given by

$$\frac{\partial}{\partial \mathbf{v}} R(\mathbf{v}) = h_t h_X^n \mathbf{B}^\top \mathbf{B} \mathbf{v} \quad \text{and} \quad \frac{\partial^2}{\partial \mathbf{v}^2} R(\mathbf{v}) = h_t h_X^n \mathbf{B}^\top \mathbf{B}.$$

In order to obtain an efficient iterative second-order method for solving (23), one requires an approximation of the Hessian $\mathbf{H} \in \mathbb{R}^{N \times N}$ that balances the following tradeoff. Ideally, it is reasonably efficient to compute, consumes limited memory (sparsity is desired), and has sufficient structure so that preconditioning can be used. However, each iteration of the Gauss–Newton method should also provide a suitable descent direction. For these reasons, we approximate the Hessian by

$$\begin{aligned} \mathbf{H}(\mathbf{v}) = \frac{\partial^2}{\partial \mathbf{v}^2} J_{\gamma, \mathbf{g}}(\mathbf{v}) \approx & \frac{\partial}{\partial \mathbf{v}} \mathbf{f}(\mathbf{v})^\top \frac{\partial}{\partial \mathbf{f}} \mathbf{K}(\mathbf{f}(\mathbf{v}))^\top \frac{\partial^2}{\partial \mathbf{x}^2} D(\mathbf{K}(\mathbf{f}(\mathbf{v})), \mathbf{g}) \frac{\partial}{\partial \mathbf{f}} \mathbf{K}(\mathbf{f}(\mathbf{v})) \frac{\partial}{\partial \mathbf{v}} \mathbf{f}(\mathbf{v}) \\ & + \gamma h_t h_X^n \mathbf{B}^\top \mathbf{B} + \epsilon \mathbf{I}_N, \end{aligned}$$

where $\epsilon > 0$ ensures positive semidefiniteness. For simplicity, the term involving $\partial^2 \mathbf{f}(\mathbf{v})/\partial \mathbf{v}^2$ is omitted and, regardless of the chosen distance, we use the second derivative in (26) as an approximation of $\partial^2 D(\mathbf{x}, \mathbf{y})/\partial \mathbf{x}^2$. In our numerical experiments we found that this choice works well for the problem considered in Sec. 5.

It remains to discuss the derivative of the solution map. For the transport equation, the application of the chain rule to (21) yields

$$\frac{\partial}{\partial \mathbf{v}} \mathbf{f}(\mathbf{v}) = \nabla_x \mathbf{f}_0(\mathbf{X}_{\mathbf{v}}(1, 0, \mathbf{x}_c)) \frac{\partial}{\partial \mathbf{v}} \mathbf{X}_{\mathbf{v}}(1, 0, \mathbf{x}_c),$$

where $\nabla_x \mathbf{f}_0$ denotes the gradient of the interpolation of the template image and $\partial \mathbf{X}_{\mathbf{v}}/\partial \mathbf{v}$ is the derivative of the endpoints of the characteristic curves with respect to the velocities, see below. Similarly, for the solution map (22) that corresponds to the continuity equation, we obtain

$$\frac{\partial}{\partial \mathbf{v}} \mathbf{f}(\mathbf{v}) = \frac{\partial}{\partial \mathbf{X}_{\mathbf{v}}} (\mathbf{F}(\mathbf{X}_{\mathbf{v}}(0, 1, \mathbf{x}_c)) \mathbf{f}_0(\mathbf{x}_c)) \frac{\partial}{\partial \mathbf{v}} \mathbf{X}_{\mathbf{v}}(0, 1, \mathbf{x}_c).$$

Here, $\partial \mathbf{F}/\partial \mathbf{X}_{\mathbf{v}}$ is the derivative of the pushforward matrix with respect to the endpoints of the characteristics, again see [35, Sec. 3.1].

If explicit time stepping methods are used to solve the ODE (6), the partial derivative $\partial \mathbf{X}_{\mathbf{v}}/\partial \mathbf{v}$ can be computed recursively. For example, for the forward Euler approach in (20) it is given by

$$\begin{aligned} \frac{\partial}{\partial \mathbf{v}} \mathbf{X}_{\mathbf{v}}(0, t_{k+1}, \mathbf{x}_c) &= \frac{\partial}{\partial \mathbf{v}} \mathbf{X}_{\mathbf{v}}(0, t_k, \mathbf{x}_c) + \Delta t \frac{\partial}{\partial \mathbf{v}} \mathbf{I}(\mathbf{v}, t_k, \mathbf{X}_{\mathbf{v}}(0, t_k, \mathbf{x}_c)) \\ &\quad + \Delta t \frac{\partial}{\partial \mathbf{X}_{\mathbf{v}}} \mathbf{I}(\mathbf{v}, t_k, \mathbf{X}_{\mathbf{v}}(0, t_k, \mathbf{x}_c)) \frac{\partial}{\partial \mathbf{v}} \mathbf{X}_{\mathbf{v}}(0, t_k, \mathbf{x}_c), \end{aligned}$$

for all $k = 0, 1, \dots, N_t - 1$, with $\partial \mathbf{I}/\partial \mathbf{v}$ and $\partial \mathbf{I}/\partial \mathbf{X}_{\mathbf{v}}$ being the derivatives of the interpolation schemes with respect to the velocities and with respect to the endpoints of the characteristics, respectively. We refer to [39, Chap. 3.5] for details. The case where characteristics are computed backwards in time can be handled similarly.

In order to solve the finite-dimensional minimisation problem (23), we apply an inexact Gauss–Newton–Krylov method, which proceeds as follows. Given an initial guess $\mathbf{v}^{(0)} = \mathbf{0}$, we update the velocities in each iteration $i = 0, 1, \dots$ by $\mathbf{v}^{(i+1)} = \mathbf{v}^{(i)} + \mu \delta \mathbf{v}$ until a termination criterion is satisfied. Here, $\mu \in \mathbb{R}$ denotes a step size that is determined via Armijo line search and $\delta \mathbf{v} \in \mathbb{R}^N$ is the solution to the linear system

$$\mathbf{H}(\mathbf{v}^{(i)}) \delta \mathbf{v} = - \frac{\partial}{\partial \mathbf{v}} J_{\gamma, \mathbf{g}}(\mathbf{v}^{(i)}). \quad (27)$$

For details on the stopping criteria and the line search we refer to [39, Chap. 6.3.3]. We solve the system (27) approximately by means of a preconditioned conjugate gradient (PCG) method, which can be implemented matrix-free whenever the derivative of \mathbf{K} and its adjoint can be computed matrix-free. See [35, Sec. 3.2] for further details on the preconditioning.

Due to the non-convexity of (17) and to speed up computation, we use a multilevel strategy in order to reduce the risk of ending up in a local minimum, see [27]. On each level, we use a subsampled version of the velocities that were computed on the previous, more coarser, discretisation as initial guess.

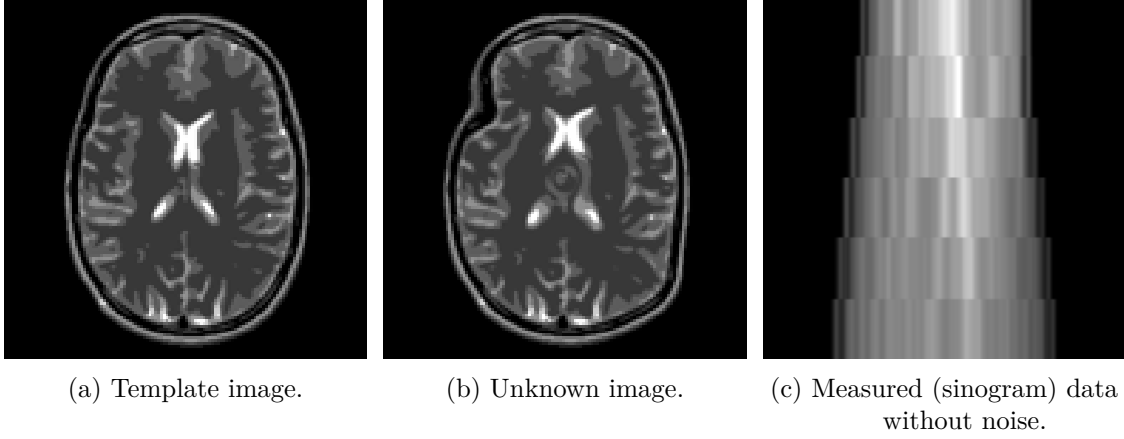


Figure 1: Synthetic example based on an artificial brain image [25] that has been deformed manually. We generated six Radon transform measurements that correspond to six equally spaced angles from the interval $[0, 60]$ degrees.

While standard image registration typically uses resampling of the template and the target image [39, Chap. 3.7], the approach described here requires multilevel versions of the operator \mathbf{K} together with a suitable method for resampling the measurements \mathbf{g} . We stress that, if these are not available, optimisation can as well just be performed on the finest discretisation level.

In the following, we assume that \mathbf{K} is a discretisation of the Radon transform [41], which is a linear operator, and outline a suitable procedure for creating multilevel versions of the operator and the measured data. The former is easily achieved with a computational backend such as ASTRA [54, 55], which allows to explicitly specify the number of grid cells used to discretise the measurement geometry. For the sake of simplicity, we restrict the presentation here to the case where $n = 2$, i.e. $\Omega \subset \mathbb{R}^2$, and \mathbf{K} is linear.

Let us assume that the number of grid cells used to discretise Ω at the finest level is $m = 2^\ell$, $\ell \in \mathbb{N}$. In our experiments, we set the number of grid cells of the one-dimensional measurement domain $(0, 1)$ at the current level $k \leq \ell$ to $q^{(k)} = 1.5 \cdot 2^{(k)}$ and set the length of each cell to $h_Y^{(k)} = 1/q^{(k)}$. Then, a multilevel representation of each measurement \mathbf{g}_i , $i \leq p$, at cell-centred grid points $\mathbf{y}_j = (j - 1/2)h_Y^{(k-1)}$ is given by

$$\mathbf{g}_i^{(k-1)}(\mathbf{y}_j) := \left(\mathbf{g}_i^{(k)}(\mathbf{y}_j) + \mathbf{g}_i^{(k)}(\mathbf{y}_j + h_Y^{(k)}) \right) / 4,$$

where the denominator arises from averaging over two neighbouring grid points and dividing the edge length of the imaging domain Ω in each coordinate direction in half. The approach can be extended to higher dimensions.

5 Numerical examples

In our numerical experiments we use the Radon transform [41] as operator. Other choices are possible and, assuming that one has access to a suitable resampling procedure for the measured data, the multilevel strategy can be applied as well. The aim here is to investigate

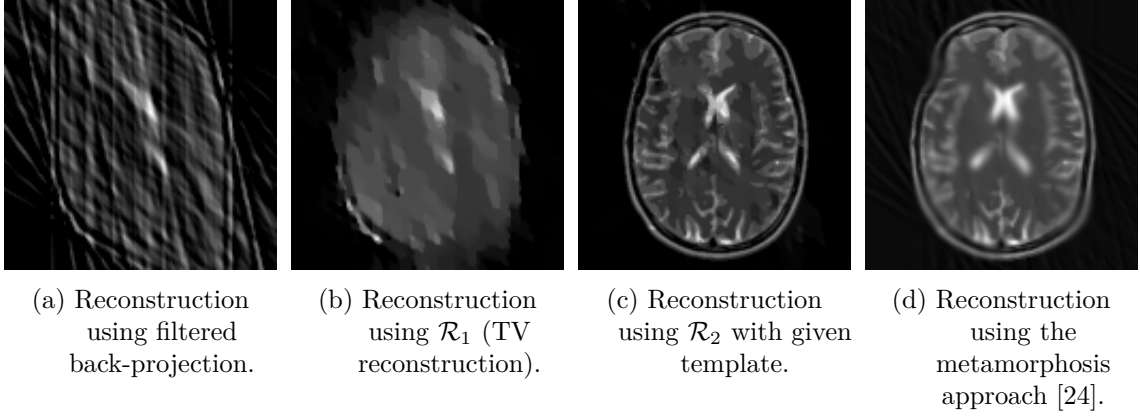


Figure 2: Comparison of different reconstruction models applied to an artificial brain image [39] that has been deformed manually. We generated six measurements that correspond to six equally spaced angles from the interval $[0, 60]$ degrees.

the reconstruction quality with different regularisation functionals, distances, and noise levels for both PDE constraints. We show synthetic examples for the settings $n = 2$ and $n = 3$, and non-synthetic examples for $n = 2$ using real X-ray tomography data. In the synthetic case, all shown reconstructions were computed from measurements taken from at most 10 directions (i.e. angles) sampled from intervals within $[0, 180]$ degrees.

All computations were performed using an Intel Xeon E5-2630 v4 2.2 GHz server equipped with 128 GB RAM and an NVIDIA Quadro P6000 GPU featuring 24 GB of memory. The GPU was only used for computing the Radon transform of 3D volumes.

Before we proceed, we give a brief idea of suitable parameter choices. For the multilevel approach we used in each synthetic example 32×32 pixels at the coarsest level and 128×128 pixels at the finest level, i.e. $\ell = 7$. The size of the reconstructed images in the nonsynthetic examples was 128×128 . Again, three levels were used. In the synthetic 3D example the reconstructed volume was $32 \times 32 \times 32$ and the coarsest level was $8 \times 8 \times 8$.

We used time dependent velocity fields with only one time step, i.e. $n_t = 1$, since this keeps the computational cost reasonable and sufficed for our examples. The characteristics were computed using five Runge–Kutta steps, i.e. $N_t = 5$.

The spatial regularisation parameter depends on the chosen regularisation functional and the noise level, and was chosen in the order of 10^{-3} , 10^0 , and 10^3 for third-order, curvature, and diffusion regularisation, respectively, in the noise-free case and using the NCC-based distance. The temporal regularisation parameter is less sensitive and was chosen in the order of 10^2 . Furthermore, the parameter corresponding to the norm of $L^2(\Omega, \mathbb{R}^n)$ in (19) was set to 10^{-6} .

In our first example, we investigate different regularisation functionals with different noise levels together with the transport equation. The target is 2D Radon transform data based on a digital brain image and the template is a deformed version thereof, see Fig. 1. Since we want to focus on the behaviour of the regularisation functionals, we do not treat the continuity equation here. The data was generated using parallel beam tomography with only six equally distributed angles from the interval $[0, 60]$ degrees and was corrupted with Gaussian white noise of different levels.

Fig. 2 shows results obtained from the generated noise-free measurements using four existing methods. In Fig. 2a filtered backprojection was used. In Figs. 2b and 2c, the following two total variation regularisation-based models, see e.g. [10],

$$\min_{\mathbf{u}} \|\mathbf{K}\mathbf{u} - \mathbf{g}\|^2 + \gamma \mathcal{R}_i(\mathbf{u}),$$

with $\mathcal{R}_1(\mathbf{u}) := \text{TV}(\mathbf{u})$, $\mathcal{R}_2(\mathbf{u}) := \text{TV}(\mathbf{u} - \mathbf{f}_0)$, and $\gamma > 0$ were used. Here, $\mathcal{R}_2(\mathbf{u})$ incorporates template information. Approximate minimisers of both functionals were computed using the primal-dual hybrid gradient method [11]. For the case of filtered backprojection, the standard MATLAB implementation was used. The results in Figs. 2a to 2c highlight why more sophisticated methods, such as the proposed template-based approach, are necessary to obtain satisfying reconstructions in this setting, and illustrate the challenges when dealing with very sparse data.

As outlined in Sec. 1, one possibility is the metamorphosis approach [24]. In Fig. 2d we show a result obtained with this method using the recommended parameters. However, 200 iterations of gradient descent were performed, and the regularisation parameters were set to $\gamma = 10^{-5}$ and $\tau = 1$. Observe the change in image intensities compared to Fig. 1a and the blur in the heavily deformed regions.

In Fig. 3, we show results for the different noise levels and different regularisation functionals computed with our approach. All results were obtained using the NCC-based distance. As expected, the quality of the reconstruction gets worse for higher noise levels and, consequently, larger regularisation parameters were necessary. Since data is acquired from only six directions, the influence of the noise is very strong. Especially for the diffusive regularisation we needed to choose large regularisation parameters for higher noise levels, see Fig. 3c. Since diffusion corresponds to first-order regularisation, it is much easier to reconstruct the noise with “rough” deformations. Overall, we found that second- and third-order regularisation performed similar when appropriate regularisation parameters were chosen. Even though some theoretical results only hold for higher-order regularity, second-order regularisation seems sufficient for our use case. The computation time for the results in Fig. 3 was between 200 and 700 seconds.

In the second example, see Fig. 4, we compare the behaviour of the SSD and the NCC-based distance. The example consists of two different hands which, in addition, are rotated relative to each other. Here, the deformation is much larger than in the previous example, but still fairly regular. The data was generated similarly to the previous example, but with only five angles from the interval $[0, 75]$ degrees. Note also that the intensities of the template and target image are different (roughly by a factor of two). First, we discuss the transport equation. The intensity difference is a serious issue if we use the SSD distance, as we can see in Fig. 4e. The hand is deformed into a smaller version in order to compensate the differences. If we use the NCC-based distance instead, which can deal with such discrepancies, the result is much better from a visual point of view. The shapes are well-aligned. The resulting SSIM value is still low, which is not surprising since SSIM is not invariant with respect to intensity differences between perfectly aligned images. However, neither of the two approaches is able to remove or create any of the additional (noise) structures in the images. For the combination SSD with continuity equation, no satisfactory results could be obtained. Since no change of intensity is possible by changing the size of the hand, part of it is moved outside of the image. This behaviour could potentially be corrected if

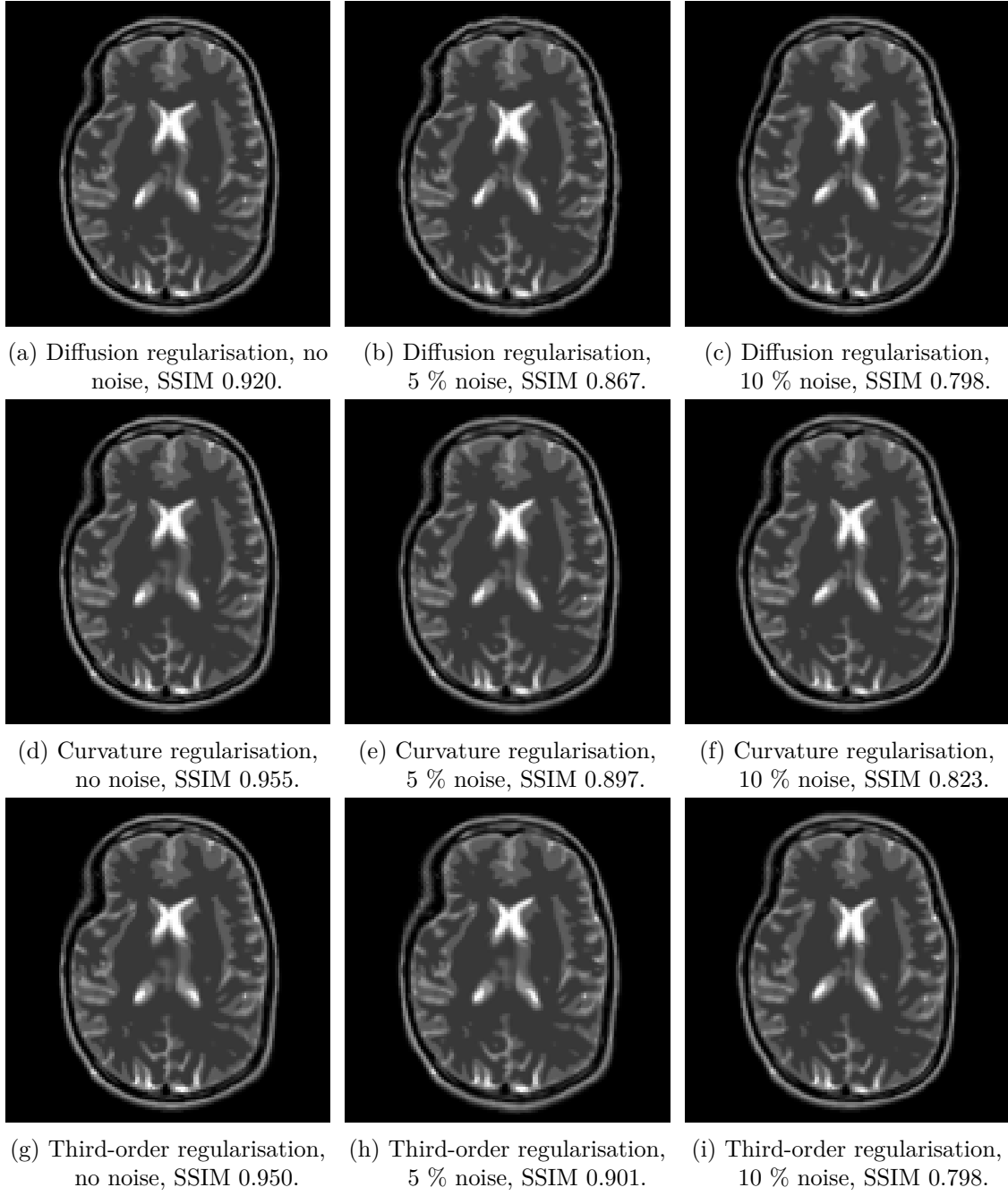


Figure 3: Reconstructions for the artificial brain image in Fig. 1 using our method and different regularisation functionals. Note that only six measurements were used. The measured data was corrupted with noise of different levels.

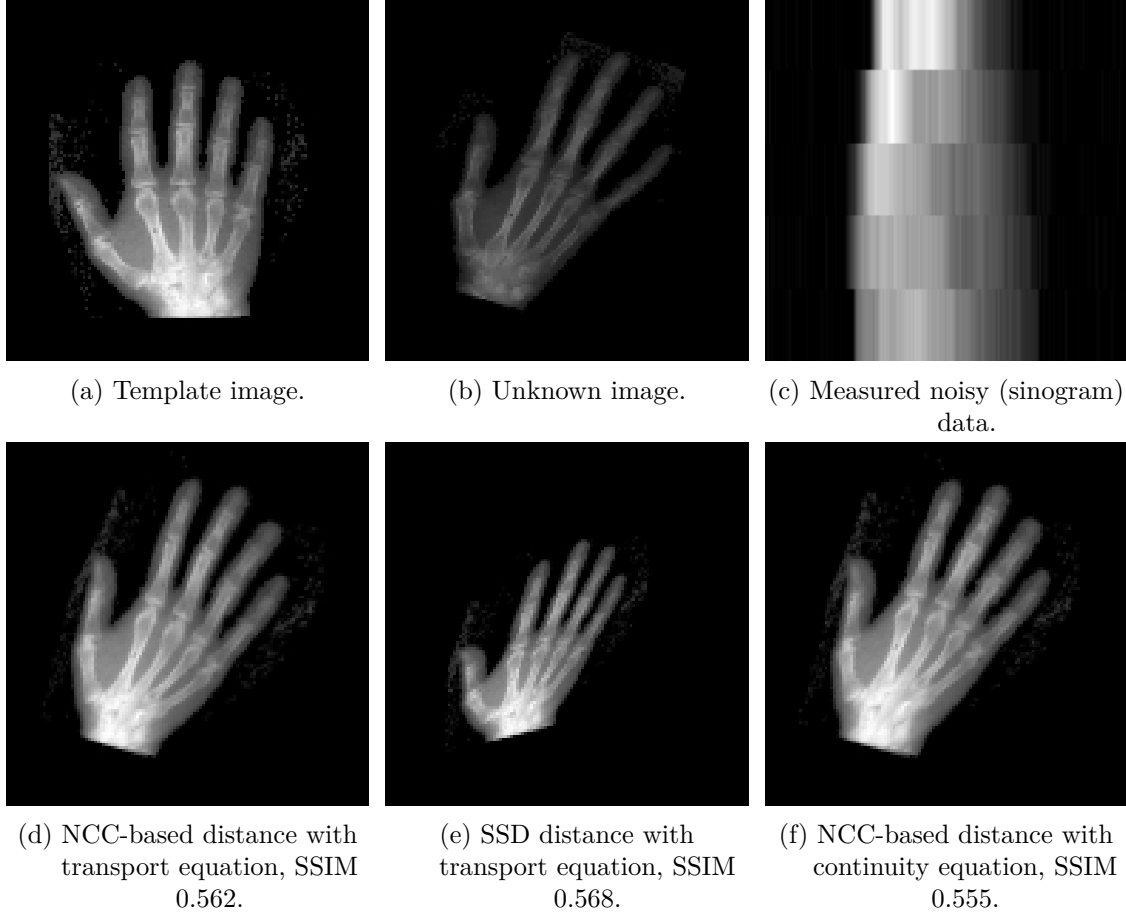


Figure 4: Reconstructions of manually deformed Hand [39] images with different image intensity levels using our method. We generated five measurements that correspond to five equally spaced angles from the interval $[0, 75]$ degrees and added five percent noise.

other boundary conditions are used in the implementation. Therefore, we do not provide an example image for this case. Using the NCC-based distance, the results look similar as for the transport equation with slightly worse SSIM value. These results suggest that the NCC-based distance is a more robust choice that avoids unnatural deformations, which would be required in the case of SSD to compensate intensity differences. In this example, the computation time was between 50 and 325 seconds.

In the next example, see Fig. 5, we compare the continuity equation with the transport equation as constraint together with the NCC-based distance. The continuity equation allows for limited change of mass along the deformation path. Since the intensity change scales with the determinant of the Jacobian, bigger changes are only possible if areas are compressed or extended a lot. In the presented example this occurs only to a mild extent. For this example, the continuity equation and the transport equation yield visually similar results with minor differences in the SSIM value. As in the previous examples, higher-order regularisation is beneficial and artefacts occur for the diffusion regularisation. The

computation time amounted to roughly 64 to 360 seconds in this example.

In Fig. 6, we created an artificial pair of images consisting of a disk to show the possibilities of intensity changes when using the continuity equation as a constraint. Both template and unknown image were constructed so that their total mass is equal. The measurements were generated as before using only five angles uniformly distributed in the interval $[0, 90]$ degrees. Furthermore, we used curvature regularisation. For the transport equation we observe that the shape is matched, but the intensity is not correct, see Fig. 6d. If we use the continuity equation instead, intensity changes are possible, which can be observed in Fig. 6e. The computation time for the two results was 90 and 500 seconds.

In order to demonstrate the practicality of our method, we computed results from non-synthetic X-ray tomography data [8, 28], which are available online.^{2,3} See Figure 7 for these two examples ('lotus' and 'walnut'). The template was generated by applying filtered back-projection to the full measurements and by subsequently deforming it. Then, this deformed templated was used in our method to compute a reconstruction from only few measurement directions. The computation time amounted to roughly 80 and 600 seconds in these examples. In both nonsynthetic examples the use of the NCC-based distance proved crucial and no satisfactory result could be obtained using SSD.

In Fig. 8, we demonstrate that our framework is also capable of reconstructing 3D volumes. Here, we used the SSD distance together with curvature regularisation and the transport equation. We applied the 3D Radon transform to obtain ten measurements from angles within $[0, 180]$. The total computation time was roughly 800 seconds.

All in all, our results demonstrate that, given a suitable template image, very reasonable reconstructions can efficiently be obtained from only a few measurements, even in the presence of noise. Moreover, our examples show that the NCC-based distance adds robustness to the approach with regard to discrepancies in the image intensities.

6 Conclusions

Overall, our numerical examples show that our implementation yields good results, as long as the deformation between template and target is fairly regular. By using the NCC-based distance, robustness with respect to intensity differences between the template and the target image can be achieved. As already mentioned in the introduction, we do not follow the metamorphosis approach, since there is too much flexibility in the model and the source term is very likely to reproduce noise and artefacts if the data is too limited. It is left for further research to investigate possible adaptations of the model that allow for the appearance of new objects or structures in the reconstruction without reproducing noise or artefacts. Possibly, the results of our method can be used as better template for other algorithms that require template information. Finally, note that due to the great flexibility of the FAIR library, it is also possible to use a great variety of regularisation functionals for the velocities and other distances, see [39, Chaps. 7 and 8]. Additionally, our implementation is not necessarily restricted to the Radon transform and essentially every (continuous) operator can be used. The multilevel approach can be applied as long as a meaningful resampling procedure for the operator and the measured data can be provided.

²<https://doi.org/10.5281/zenodo.1254204>

³<https://doi.org/10.5281/zenodo.1254206>

Acknowledgments

Lukas F. Lang and Carola-Bibiane Schönlieb acknowledge support from the Leverhulme Trust project “Breaking the non-convexity barrier”, the EPSRC grant EP/M00483X/1, the EPSRC Centre Nr. EP/N014588/1, the RISE projects ChiPS and NoMADS, the Cantab Capital Institute for the Mathematics of Information, and the Alan Turing Institute. Sebastian Neumayer is funded by the German Research Foundation (DFG) within the Research Training Group 1932, project area P3. Ozan Öktem is supported by the Swedish Foundation of Strategic Research, grant AM13-0049. We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Quadro P6000 GPU used for this research.

References

- [1] H. W. Alt. *Linear Functional Analysis: An Application-Oriented Introduction*. Universitext. Springer-Verlag, London, 2016, pages xii+435.
- [2] R. Andreev, O. Scherzer, and W. Zulehner. Simultaneous optical flow and source estimation: space-time discretization and preconditioning. *Appl. Numer. Math.*, 96:72–81, 2015.
- [3] J. Ashburner. A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95–113, 2007.
- [4] M. F. Beg, M. I. Miller, A. Trounev, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vision*, 61(2):139–157, 2005.
- [5] A. Borzi, K. Ito, and K. Kunisch. An optimal control approach to optical flow computation. *Int. J. Numer. Methods Fluids*, 40(1-2):231–240, 2002.
- [6] A. Borzi, K. Ito, and K. Kunisch. Optimal control formulation for determining optical flow. *SIAM J. Sci. Comput.*, 24(3):818–847, 2003.
- [7] A. Borzi and V. Schulz. *Computational Optimization of Systems Governed by Partial Differential Equations*, volume 8 of *Computational Science & Engineering*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2012. xx+282.
- [8] T. A. Bubba, A. Hauptmann, S. Huotari, J. Rimpeläinen, and S. Siltanen. Tomographic X-ray data of a lotus root filled with attenuating objects. *arXiv*, arXiv:1609.07299, 2016.
- [9] M. Burger, H. Dirks, and C.-B. Schönlieb. A variational model for joint motion estimation and image reconstruction. *SIAM J. Imaging Sciences*, 11(1):94–128, 2018.
- [10] E. J. Candès, J. K. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, 2006.
- [11] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision*, 40(1):120–145, 2011.
- [12] C. Chen and O. Öktem. Indirect image registration with large diffeomorphic deformations. *SIAM J. Imaging Sciences*, 11(1):575–617, 2018.

- [13] K. Chen and D. A. Lorenz. Image sequence interpolation using optimal control. *J. Math. Imaging Vision*, 41(3):222–238, 2011.
- [14] A. Chertock and A. Kurganov. On a practical implementation of particle methods. *Appl. Numer. Math.*, 56:1418–1431, 1999.
- [15] G. Crippa. *The flow associated to weakly differentiable vector fields*. PhD thesis, Classe di Scienze Matematiche, Fisiche e Naturali, Scuola Normale Superiore di Pisa / Institut für Mathematik, Universität Zürich, 2007.
- [16] J. C. De los Reyes. *Numerical PDE-Constrained Optimization*. Springer International Publishing, 2015.
- [17] R. J. DiPerna and P.-L. Lions. Ordinary differential equations, transport theory and Sobolev spaces. *Invent. Math.*, 98(3):511–547, 1989.
- [18] P. Dupuis, U. Grenander, and M.I. Miller. Variational problems on flows of diffeomorphisms for image matching. *Quart. Appl. Math.*, 56:587–600, 1998.
- [19] A. Effland. *Discrete Riemannian Calculus and A Posteriori Error Control on Shape Spaces*. Dissertation, University of Bonn, 2017.
- [20] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*, number 375 in Mathematics and its Applications. Kluwer Academic Publishers Group, Dordrecht, 1996. viii+321.
- [21] L. C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.
- [22] J. Friel. Sparse regularization in limited angle tomography. *Appl. Comput. Harmon. Anal.*, 34(1):117–141, 2013.
- [23] D. Gilbarg and N. S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001, pages xiv+517. Reprint of the 1998 edition.
- [24] B. Gris, C. Chen, and O. Öktem. Image reconstruction through metamorphosis. Technical report, 2018. arXiv: [1806.01225 \[cs.CV\]](#).
- [25] M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, and M. Unser. Realistic analytical phantoms for parallel magnetic resonance imaging. *IEEE Trans. Med. Imag.*, 31(3):626–636, 2012.
- [26] M. D. Gunzburger. *Perspectives in Flow Control and Optimization*. Advances in Design and Control. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2003. xiv+261.
- [27] E. Haber and J. Modersitzki. A multilevel method for image registration. *SIAM J. Sci. Comput.*, 27(5):1594–1607, 2006.
- [28] K. Hämäläinen, L. Harhanen, A. Kallonen, A. Kujanpää, E. Niemi, and S. Siltanen. Tomographic X-ray data of walnut. *arXiv*, arXiv:1502.04064, 2015.
- [29] G. L. Hart, C. Zach, and M. Niethammer. An optimal control approach for deformable registration. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, volume 2, pages 1223–1227. IEEE, 2009.

- [30] R. Herzog and K. Kunisch. Algorithms for PDE-constrained optimization. *GAMM-Mitt.*, 33(2):163–176, 2010.
- [31] J. Hinkle, M. Szegedi, B. Wang, B. Salter, and S. Joshi. 4D CT image reconstruction with diffeomorphic motion model. *Med. Image Anal.*, 16(6):1307–1316, 2012.
- [32] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE Constraints*, volume 23 of *Mathematical Modelling: Theory and Applications*. Springer, New York, 2009, pages xxi+270.
- [33] Y. Hong, S. Joshi, M. Sanchez, M. Styner, and M. Niethammer. Metamorphic geodesic regression. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012*. Volume 7512, Lecture Notes in Computer Science, pages 197–205. Springer Berlin Heidelberg, 2012.
- [34] J. Maas, M. Rumpf, C.-B. Schönlieb, and S. Simon. A generalized model for optimal transport of images including dissipation and density modulation. *ESAIM: M2NA*, 49(6):1745–1769, 2015.
- [35] A. Mang and L. Ruthotto. A Lagrangian Gauss–Newton–Krylov solver for mass- and intensity-preserving diffeomorphic image registration. *SIAM J. Sci. Comput.*, 39(5):B860–B885, 2017.
- [36] M. I. Miller, A. Trounev, and L. Younes. Geodesic shooting for computational anatomy. *J. Math. Imaging Vision*, 24(2):209–228, 2006.
- [37] M. I. Miller, A. Trounev, and L. Younes. Hamiltonian systems and optimal control in computational anatomy: 100 years since D’Arcy Thompson. *Annu. Rev. Biomed. Eng.*, 17(1):447–509, 2015.
- [38] M. I. Miller and L. Younes. Group actions, homeomorphisms, and matching: a general framework. *Int. J. Comput. Vision*, 41(1/2):61–84, 2001.
- [39] J. Modersitzki. *FAIR: Flexible Algorithms for Image Registration*, volume 6 of *Fundamentals of Algorithms*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009.
- [40] J. Modersitzki. *Numerical Methods for Image Registration*. Oxford University Press, New York, 2003.
- [41] F. Natterer. *The Mathematics of Computerized Tomography*, volume 32 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. xviii+222.
- [42] S. Neumayer, J. Persch, and G. Steidl. Morphing of manifold-valued images inspired by discrete geodesics in image spaces. *SIAM J. Imaging Sciences*, 11(3):1898–1930, 2018.
- [43] S. Neumayer, J. Persch, and G. Steidl. Regularization of inverse problems via time discrete geodesics in image spaces. Technical report, 2018. arXiv: [1805.06362 \[math.NA\]](https://arxiv.org/abs/1805.06362).
- [44] M. Niethammer, G. L. Hart, and C. Zach. An optimal control approach for the registration of image time-series. In *Proceedings of the 48th IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 262–270. IEEE, 2009.

- [45] O. Öktem, C. Chen, N. O. Domanic, P. Ravikumar, and C. Bajaj. Shape-based image reconstruction using linearized deformations. *Inverse Probl.*, 33(3):035004, 2017.
- [46] C. L. Richardson and L. Younes. Metamorphosis of images in reproducing kernel Hilbert spaces. *Adv. Comput. Math.*, 42(3):573–603, 2015.
- [47] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1–4):259–268, 1992.
- [48] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*, number 167 in Applied Mathematical Sciences. Springer, New York, 2009.
- [49] N. Singh, J. Hinkle, S. Joshi, and P. T. Fletcher. A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction. In *Proceedings of the 10th International Symposium on Biomedical Imaging*, pages 127–142. IEEE, 2013.
- [50] A. Trounev. Diffeomorphisms groups and pattern matching in image analysis. *Int. J. Comput. Vision*, 28(3):213–221, 1998.
- [51] A. Trounev and L. Younes. Local geometry of deformable templates. *SIAM J. Math. Anal.*, 37(1):17–59, 2005.
- [52] A. Trounev and L. Younes. Metamorphoses through Lie group action. *Found. Comput. Math.*, 5(2):173–198, 2005.
- [53] A. Trounev and L. Younes. Shape spaces. In *Handbook of Mathematical Methods in Imaging*, pages 1759–1817. Springer New York, 2015.
- [54] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabrovolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers. Fast and flexible X-ray tomography using the ASTRA toolbox. *Opt. Express*, 24(22):25129, 2016.
- [55] W. van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers. The ASTRA toolbox: a platform for advanced algorithm development in electron tomography. *Ultramicrosc.*, 157:35–47, 2015.
- [56] F.-X. Vialard, L. Risser, D. Rueckert, and C. J. Cotter. Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation. *Int. J. Comput. Vision*, 97(2):229–241, 2011.
- [57] X. Yang, R. Kwitt, M. Styner, and M. Niethammer. Quicksilver: fast predictive image registration – a deep learning approach. *NeuroImage*, 158:378–396, 2017.
- [58] L. Younes. *Shapes and Diffeomorphisms*, volume 171 of *Applied Mathematical Sciences*. Springer-Verlag Berlin, 2010, pages xviii+434.

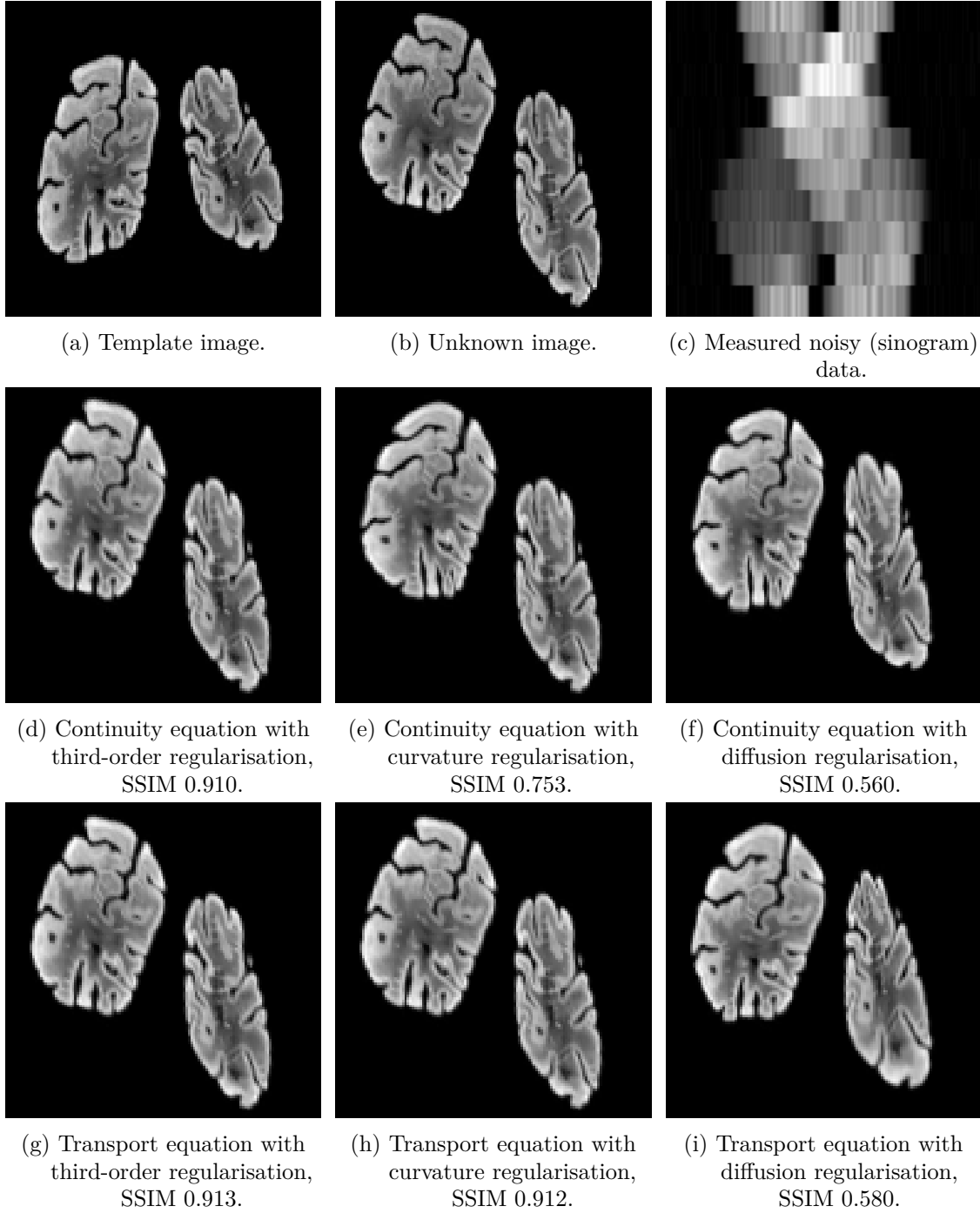


Figure 5: Reconstructions for the HNSP [39] image using our approach, different regularisation functionals, and different PDE constraints. Here, ten measurements corresponding to ten angles equally distributed in the interval $[0, 180]$ degrees were taken. The measured data was corrupted with five percent noise.

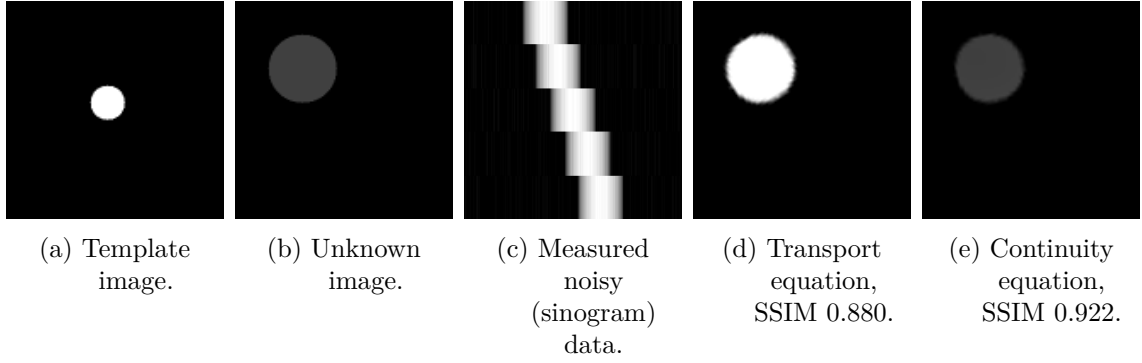


Figure 6: Reconstructions of an image showing a disk obtained with our method. Five measurements were taken at directions corresponding to five angles equally distributed in $[0, 90]$ degrees. As before, five percent noise was added.

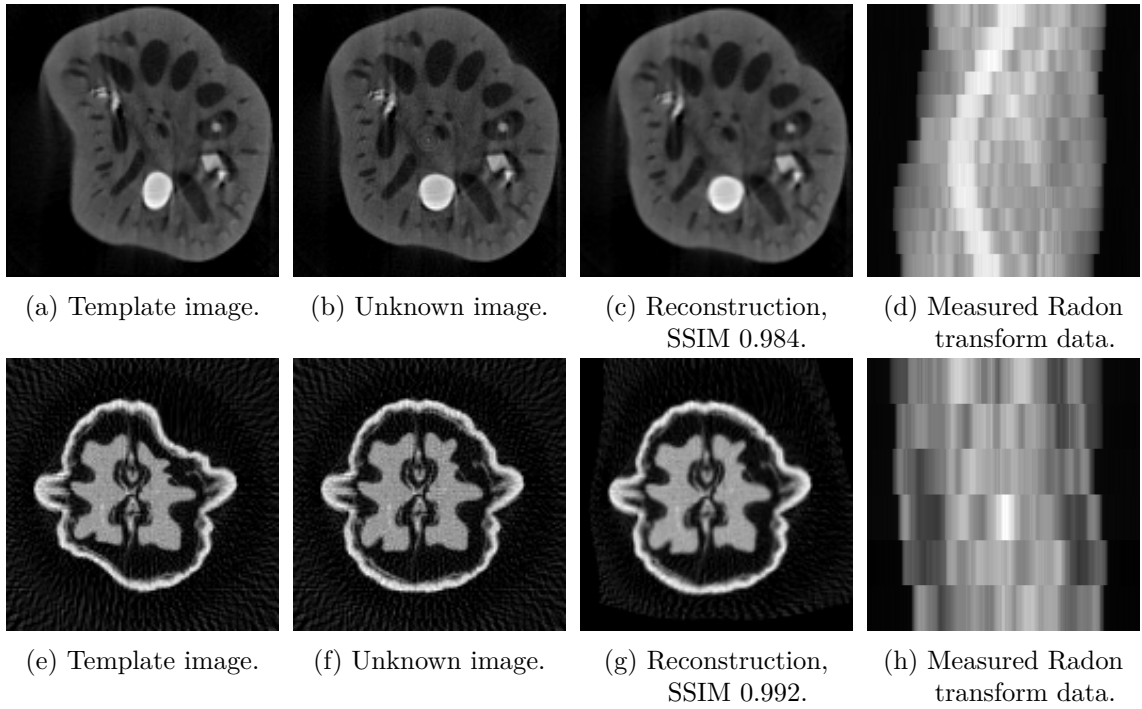


Figure 7: Reconstructions based on nonsynthetic X-ray tomographic measurements [8, 28] computed with our method using the transport equation together with curvature regularisation. Measurements from twelve and six directions with angles in $[0, 180]$ degrees were used.

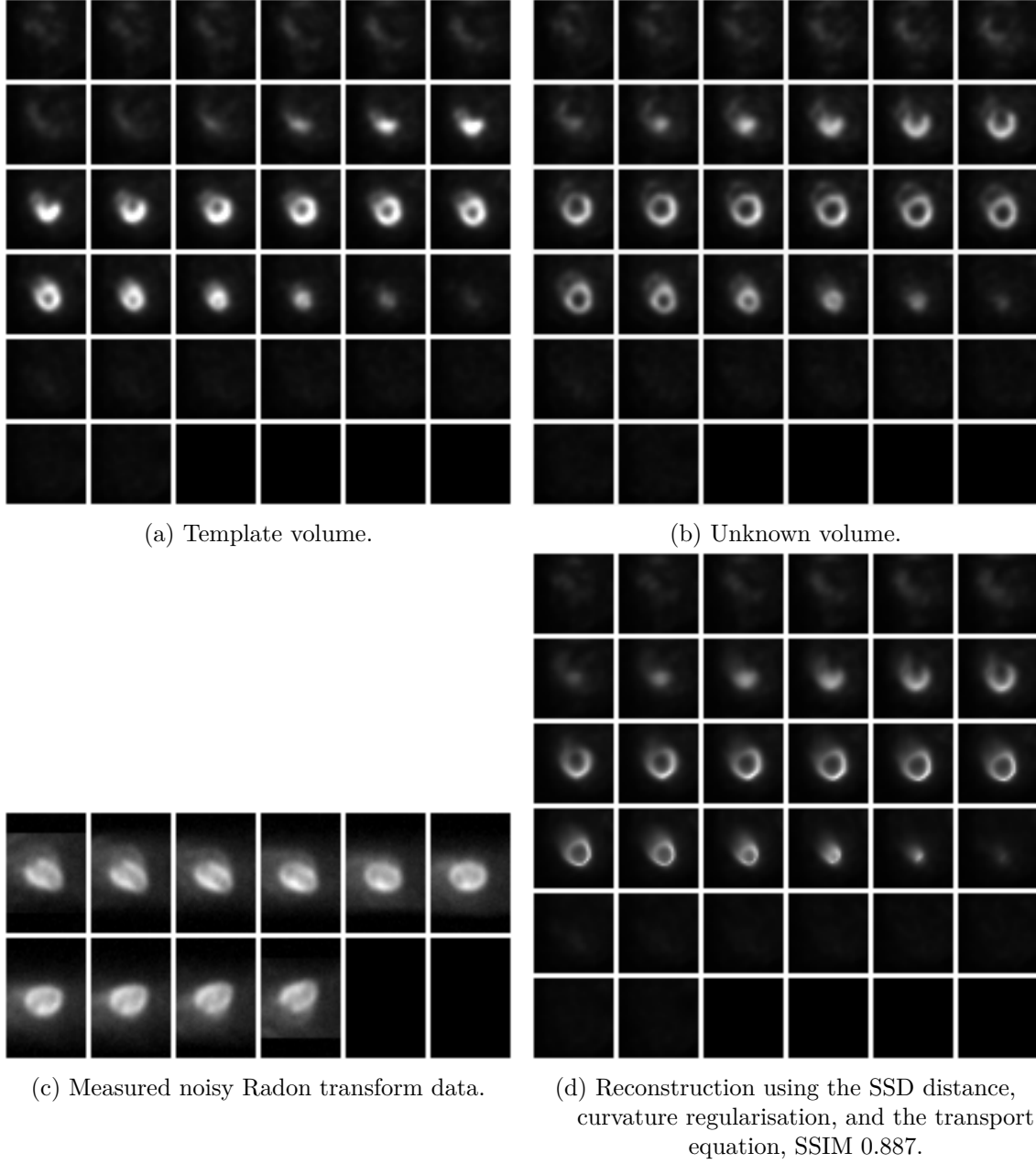


Figure 8: Reconstruction of a 3D volume (‘mice3D’, see [39]) using our method. In Fig. 8a, Fig. 8b, and Fig. 8d, slices (left to right, top to bottom) of each volume along the third coordinate direction are shown. In Fig. 8c, slices of the 3D Radon transform measurements are shown. Each slice corresponds to one measurements direction. In total, only ten measurements were taken at angles equally distributed in $[0, 180]$ degrees. As before, five percent noise was added.